Parallel Excitatory and Inhibitory Neural Circuit Pathways Underlie Reward-Based Phasic Neural Responses

Title Page

Title： Parallel Excitatory and Inhibitory Neural Circuit Pathways Underlie Reward Based Phasic Neural responses

Abbreviated title： Neural circuit mechanism of multiple reward based learned responses

Author names and affiliations

Huanyuan Zhou[1], KongFatt Wong-Lin[2], Da-Hui Wang[1]

1.School of Systems Science and National Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, 100875 China

2. Intelligent Systems Research Centre, School of Computing, Engineering, and Intelligent Systems, University of Ulster, Magee Campus, Northland Road, L'Derry BT48 7JL, UK

Corresponding author should be addressed: wangdh@bnu.edu.cn

Conflict of Interest: The authors declared that they have no conflicts of interest to this work.

Acknowledgements:

DHW was supported by NSFC under Grant No.31271169, 31671077 and the Fundamental Research Funds for Central University. KFW-L was supported by the Northern Ireland Functional Brain Mapping Project (1303/101154803), supported by Invest NI and the University of Ulster, BBSRC (BB/P003427/1), and COST Action Open Multiscale Systems Medicine (OpenMultiMed) supported by COST (European Cooperation in Science and Technology). DHW and KFW-L were also supported by The Royal Society–NSFC International Exchanges Scheme–Cost Share Programme (31511130066, IE141307).

Abstract

Phasic activity of dopaminergic (DA) neurons in the ventral tegmental area or substantia nigra compacta (VTA/SNc) has been suggested to encode reward prediction error signal for reinforcement learning. Recent studies have shown that the lateral habenula (LHb) neurons exhibit similar response, but for nonrewarding or punishment signals. Hence the transient signalling role of LHb neurons is opposite that of DA neurons, and also that of several other brain nuclei such as the border region of the globus pallidus internal segment (GPb) and the rostral medial tegmentum (RMTg). Previous theoretical models have investigated the neural circuit mechanism underlying reward-based phasic activity of DA neurons, but the feasibility of a larger neural circuit model to account for the observed reward based phasic activity in other brain nuclei such as the LHb has yet to be shown. Here we propose a large-scale neural circuit model and show that parallel excitatory and inhibitory pathways underlie the learned neural responses across multiple brain regions. Specifically, the model can account for the phasic neural activity observed in the GPb, LHb, RMTg and the VTA/SNc. Based on sensitivity analysis, the model is found to be robust against changes in the overall neural connectivity strength. The model also predicts that striosome plays a key role in the phasic activity of VTA/SNc and LHb neurons by encoding previous and expected rewards. Taken together, our model identifies the important role of parallel neural circuit pathways in accounting for phasic activity across multiple brain areas during reward and punishment processing.

Introduction

The ability to adapt to uncertainty is critical for survival and key to wellbeing. To investigate the underlying neural correlates and mechanisms, many experimental studies and computational studies using stochastic scheduling of reward have been carried out (Schultz et al., 1997; Fiorillo et al., 2003; Kuhnen and Knutson, 2005; McCoy and Platt, 2005; Matsumoto and Hikosaka, 2007; Hong and Hikosaka, 2008; van Duuren et al., 2009; Hong et al., 2011; Monosov and Hikosaka, 2013). Experimental studies have demonstrated that dopaminergic (DA) neurons in the ventral tegmental area or substantia nigra compacta (VTA/SNc) and the lateral habenual (LHb) play important roles in encoding uncertainty of reward and punishment (Schultz et al., 1997; Matsumoto and Hikosaka, 2007).

As illustrated schematically in Figure 1 (top row), given some unexpected reward (the presence of an unconditioned stimulus US such as food), DA (LHb) neurons exhibit a phasic peak (dip) upon the presence of the US (Schultz et al., 1997; Matsumoto and Hikosaka, 2007). After several trials of learning in the presence of a cue/stimulus, conditioning takes place. The (expected) conditioned cue/stimulus (CS) becomes associated with reward, and the DA (LHb) neurons exhibit a phasic peak (dip) in activity upon the onset of the CS (Figure 1, second row) (Schultz et al., 1997; Matsumoto and Hikosaka, 2007). Note that the DA and LHb neurons now do not respond to the unconditioned stimulus (US) with a rewarding outcome (Schultz et al.,

1997; Matsumoto and Hikosaka, 2007). One can view of this as post reinforcement learning – the agent has learned to completely associate the cue/stimulus CS with the US (e.g. an auditory tone with food), and the latter is no longer needed for further learning. However, if there is an omission of reward (e.g. absence of food), there is an additional dip (peak) in activity for the DA (LHb) neurons (Figure 1, third row) (Schultz et al., 1997; Matsumoto and Hikosaka, 2007).

Instead of the unexpected rewarding outcome US, if we now replace it with an unexpected nonrewarding or aversive stimulus US (e.g. no food or mild electric shock), in the initial phase of the reinforcement learning produces a phasic dip (peak) in the DA (LHb) neurons (Schultz et al., 1997; Matsumoto and Hikosaka, 2007) (Figure 1, fourth row). After learning, this information is transferred to the CS, in which the DA (LHb) neurons exhibit a phasic dip (peak) activity upon CS presentation while maintaining at baseline activity level during US (Figure 1, fifth row). When there is a sudden unexpected omission of such US or that the US has become rewarding, then there is a peak (dip) in activity of the DA (LHb) neurons (Schultz, et a.,1997; Matsumoto and Hikosaka, 2009; Bromberg-Martin et al., 2010) (Figure 1, bottom row). In summary, the phasic activities of DA and LHb neurons signal uncertainty in reward and punishment. Such signalling is also reflected in other brain regions such as the border region of the globus pallidus internal segment (GPb), the internal segment of the globus pallidus (GPi), and the rostral medial tegmentum

(RMTg) (Hong and Hikosaka, 2008; Hong et al., 2011). However, it is not clear how this information is transmitted within a larger neural circuit.

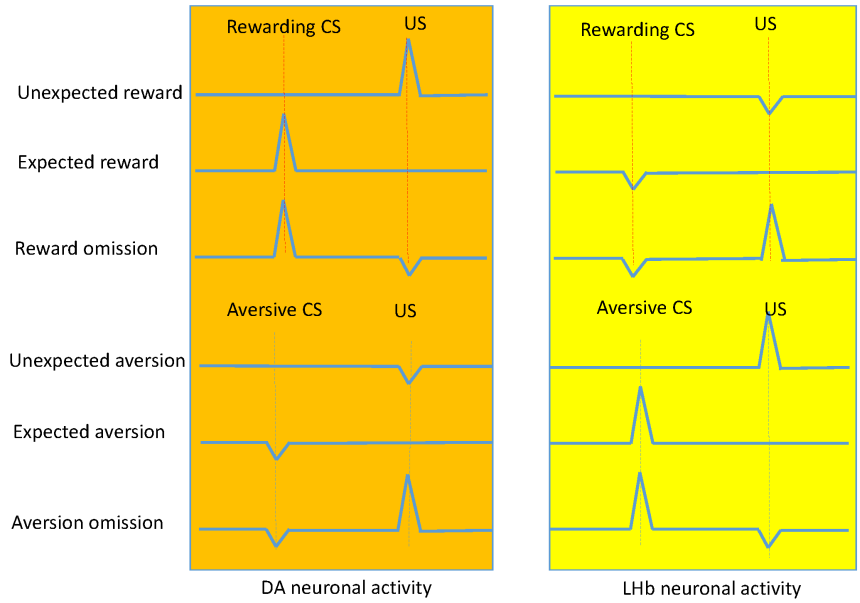

Figure 1. Schematic diagram of phasic activity of DA neurons (left orange part) and LHb neurons (right yellow part) given rewarding CS(upper) and non-rewarding/aversive CS(bottom). Each row denotes one situation of outcome

To understand the underlying computation, previous theoretical and computational studies have applied temporal difference learning (Sutton and Barto, 1981; Sutton, 1988; Schultz et al., 1997; Niv et al., 2005; Glimcher, 2011) and neural circuit modeling to understand the phasic activity of DA neurons (Brown et al., 1999; Tan and Bullock, 2008) with the basis that the phasic activity of DA neurons acts as a form of reward prediction error signal (Schultz et al., 1997). In particular, in the model by Brown et al. (1999), there are parallel pathways: one pathway from the cortex through the striosome to VTA/SNc; and another pathway from the cortex through the ventral striatum (VS) to pedunculopontine nucleus (PPTN) and VTA/SNc.

These two pathways cooperatively control the activity of DA neurons (Figure 2). However, the phasic activity of LHb neurons has not been taken into consideration yet, especially given that LHb has substantial projects to DA neurons in the VTA/SNc (Matsumoto and Hikosaka, 2007).

In this work, we propose a large-scale neural circuit model by extending the Brown et al. (1999) model to investigate the phasic activity of not only DA and LHb neurons, but also the extended parts of the network such as the GPb, GPi and RMTg. In addition to the neural circuit pathways in Brown et al. (1999) that control DA signalling (see above), our model also included pathways from the striosome and the VS to the LHb, and also one pathway from the LHb to the VTA/SNc via RMTg. These additional pathways are necessary to account for the observed phasic activity of LHb neurons (Figure 2). Further, the pathway from LHb to VTA/SNc via RMTg provides inhibition to the DA neural activity when expected reward was omitted or when there is an aversive outcome. These inter areal connectivity are constrained by currently available knowledge from physiological studies (see below for supporting evidences).

Based on simulation results, our model can account for various experimental observations of phasic activations with rewarding or nonrewarding CS, together with or without reward outcomes. Specifically, the model can account for a shift of VTA/SNc and LHb neuron responses from outcome to CS, which agrees with

experiments. In addition, the model can also account for the phasic activity of GPb and RMTg neurons, whose responses are similar to those of LHb neurons. Our model shed light on the mechanism of VTA/SNc and LHb phasic activity at the neural circuit level, with important roles from the parallel excitatory and inhibitory pathways on the learned responses, namely, that: (i) the VS-PPTN-VTA/SNc pathway excites DA, while the striosome-VTA/SNc pathway inhibits DA; (ii) VS-VP-GPb-LHb pathway inhibits LHb, while striosome-GPi-GPb-LHb pathway excites LHb; and (iii) LHb-RMTg-VTA/SNc pathway magnifies the phasic activity of VTA/SNc. The model is also rather resilient to overall changes in the inter-regional connections. Finally, our model predicts that the striosome is important since it may remember the timing of the previous reward and provide the comparison signal with the present reward.

Materials and methods

Model architecture:

Our proposed neural circuit model is schematically shown in Figure 2, which is an extended version of the model proposed by Brown et al. (1999). Namely, we included the GPb, LHb and RMTg neural populations into the model based on more recent experimental findings (Hikosaka, 2010;Hong and Hikosaka, 2008; Hong et al., 2011；Jhou, et al., 2009). The details of each part of our model are described as follows.

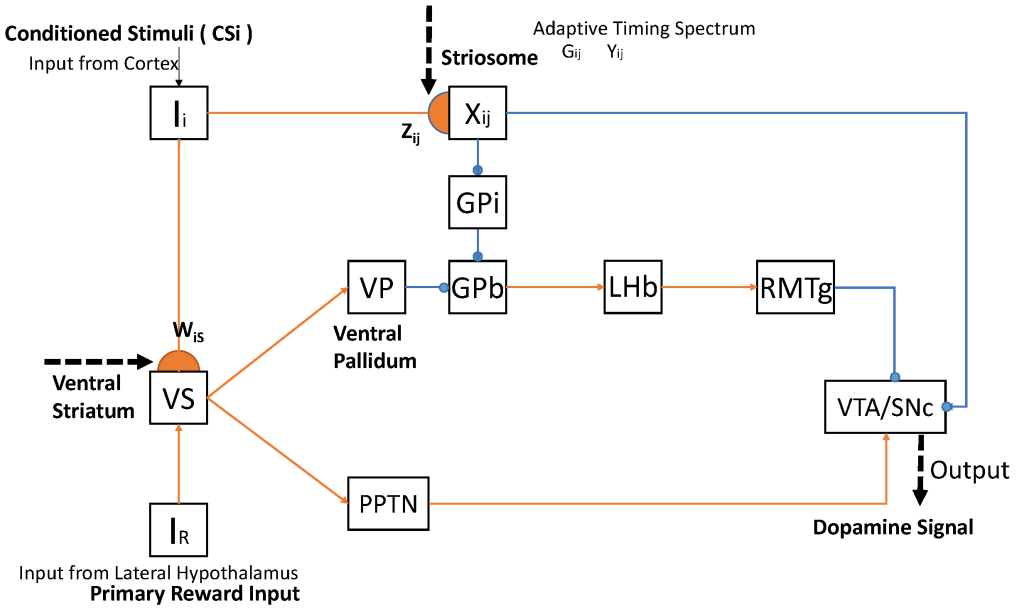

Figure 2. Model circuit. Orange arrowheads denote excitatory pathways, blue circles denote inhibitory pathways, and hemidisks denote synapses at which learning occurs. Black dashed lines denote dopaminergic signals. Evidences (Hong and Hikosaka, 2013) show that ventral striatum (VS) excite PPTN and ventral pallidum (VP). Striosome neurons project to GPi neurons which in turn project to GPb. Dopamine neurons (DA) are excited by cortical inputs (I_i) encoding conditioned stimuli and lateral hypothalamus inputs (I_R) encoding unconditioned stimuli via the path: VS-VP-GPb-LHb-RMTg-VTA/SNc and the path: VS-PPTN-VTA/SNc path. DA neurons are inhibited by I_i via the path: striosome-VTA/SNc. Note that striosome contains an adaptive spectral timing mechanism and can learn to generate lagged, adaptively timed signals (Brown et al., 1999). LHb neurons are excited by I_i via the path: striosome-GPi-GPb-LHb. LHb neurons are inhibited by I_i and I_R via the path: VS-VP-GPb-LHb.

1) LHb inhibits SNC/VTA via RMTg. Most LHb neurons are glutamatergic (Kalen et al.,1986), but experiments showed that LHb inhibits DA neurons. Firstly, in vivo recordings demonstrate that most LHb neurons are excited by a

nonreward-predicting cue and are inhibited by a reward-predicting cue when rhesus monkeys perform a visually guided saccade task (Matsumoto and Hikosaka, 2007). The phasic activity of LHb neurons is opposite to that of DA neurons in terms of responding to outcome valence–LHb (DA) neurons are excited (inhibited) by nonreward/punishment outcome/cue and inhibited (excited) by reward outcome/cue (Schultz et al., 1997; Matsumoto and Hikosaka, 2007). Secondly, LHb neurons excite earlier than the inhibition of DA neurons in unrewarded trials (Matsumoto and Hikosaka, 2007). Thirdly, stimulating LHb neurons will inhibit DA neurons (Hong and Hikosaka, 2013). The inhibition of LHb on DA neurons may arise from the direct projection from LHb neuron to inhibitory interneurons in the VTA/SNc (Brinschwitz, et al., 2010) or indirectly through some inhibitory nucleus. In fact, experiments has revealed a path from the LHb to DA neurons through RMTg and neurons in the RMTg seem to encode aversive stimuli (Jhou et al, 2009; Hikosaka, 2010). At the same time, the RMTg transmits negative reward-prediction errors signal of LHb neuron to positive reward-prediction errors signal of DA neurons (Hong et al., 2011). For simplicity, we only include the indirect path from LHb to DA neurons via GABAergic RMTg.

2) GPb excites LHb. Low intensity electrical stimulation in GPb can evoke a short latency excitatory response in LHb neurons (Hong and Hikosaka, 2013). The excitation of GPb neurons on LHb may be mediated by acetylcholine or glutamate (Hong and Hikosaka, 2008), or by disinhibition through intra-LHb interneurons

considering the complex microcircuitry within the GP (Hong and Hikosaka, 2008; Sadek et al, 2007). In addition, glutamatergic projections to LHb from entopeduncular of rat or primate's GPb neurons have been observed in experiment on non-human primates (Shabel et al., 2012; Shabel et al.,2014). In brief, there are excitatory projections from GPb to LHb and form a pathway from GPb to VTA/SNC via LHb and RMTg (Hikosaka,2010).

3) Conjectured inputs to GPb from GPi. It has been demonstrated that GPb neurons receive input from the striatum, presumably from the striosome (Rajakumar, et al.,1993). Hong and Hikosaka (2013) have observed that typical neurons in the external and internal segments of the globus pallidus (GPe and GPi) are first inhibited by striatal stimulation but GPb neurons are often (but not always) excited or disinhibited by striatal stimulations. They proposed that signals to GPb should be mediated through inhibitory axon collaterals within the striatum (Tremblay and Filion, 1989) or GPe (Sadek et al., 2007). Based on these observations, we conjecture that striosome projects to LHb through GPi.

4) VP inputs to GPb. In the Brown, et al. (1999) model, VP neurons are inhibited by the expectation of reward. However, recent experiments observe that the majority of VP neurons are excited by the expectation of a large reward (Hong and Hikosaka, 2013). Thus VP-LHb connections could possibly be inhibitory (Hong and Hikosaka, 2013). Therefore, we assume that reward-related signals are transmitted to the LHb through excitatory connections from the GPb and inhibitory connections from the VP.

5) Excitatory inputs from VS to VP and PPTN. Although VS neurons are usually identified as GABAergic and inhibit downstream neurons, Hong and Hikosaka (2013) showed that the striatal (GABAergic) neurons excite PPTN and VP neurons. The excitation by VS neurons can be mediated by substance P (Napier et al., 1995; Blomely et al.,2009). Thus, we assume that VS directly excites PPTN and VP.

Dynamical equations, input-output functions and numerical method:

We assume neuronal homogeneity within each brain regions, such that each neural population's firing rate activity within a brain region or nucleus can be dynamically described by ordinary differential equations typically with a decay term plus a term with an input-output function – firing-rate type model (Wilson and Cowan, 1976; see Mathematics and Equations section). Specifically, the neural population firing rate (output) is normalized, ranging from zero to one. The input includes constant background input to generate the spontaneous baseline firing activity for each neural population (and brain region), and synaptic terms in the form of coupling strengths to provide the interaction between different neural populations (see Mathematics and Equations). Some of the coupling strengths are subjected to change (i.e. plastic) dependent on the presence of reward (see Figure 2). Further modeling details can be obtained from original Brown et al. (1999) model. The model variables are summarized in Table 1. Parameters are adjusted to fit the observed responses of neurons. Parameter values used for simulations are given in Table 2. In all simulations,

numerical integration of the ordinary differential equations was performed with fourth-order Runge-Kutta method (Press, et al., 2007) using a custom Python code. Codes are available upon request.

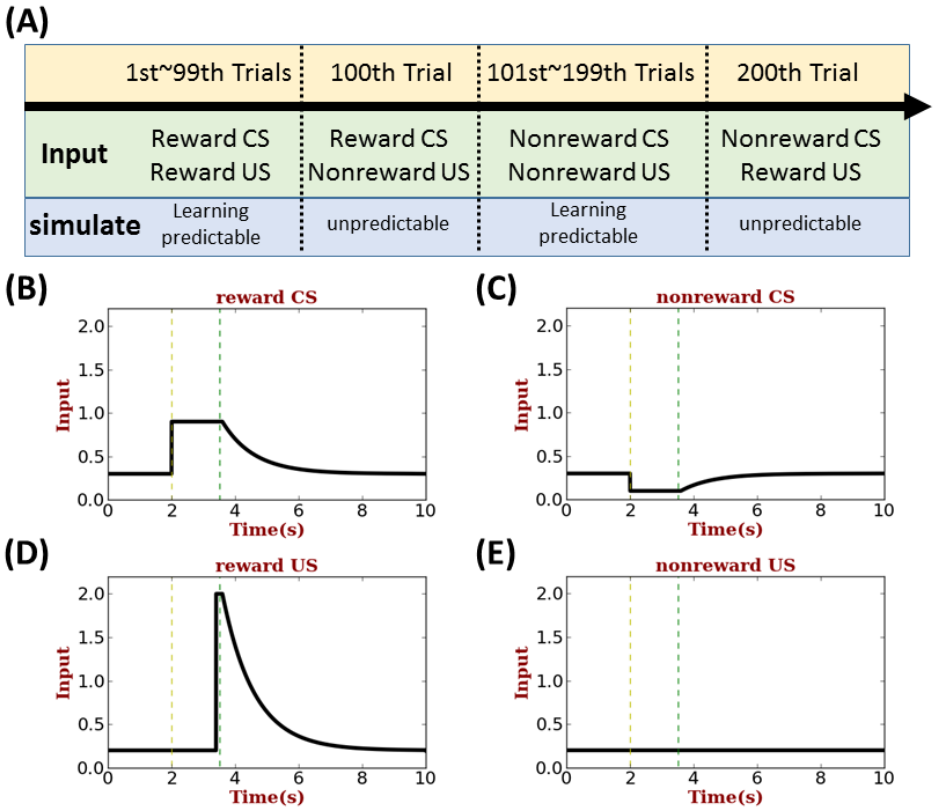

Figure 3. Model simulation protocol. (A) Different inputs are applied to simulate different conditions. We simulated a total of 200 trials. In the first 99 trials, we present reward CS input and reward US input to simulate the learning process, which associates the reward CS with the reward US. In the 100th trial, we present reward CS input but nonreward US input, thus one predicts a reward but does not receive it. In the next 99 trials, we present nonreward CS input and nonreward US input to simulate the learning process, which associates the nonreward US with the nonreward CS. In the 200th trial, we present nonreward CS input but reward US input simulating the situation where one predicts nonreward but receives it. **(B)~(E)** Different inputs. The yellow dashed line indicates the time at which CS appear (2.0 s), and the green dashed line indicates the time at which rewards are released or not (3.4 s). **(B)** Reward CS input. **(C)** Nonreward CS input. **(D)** Reward US input. **(E)** Nonreward US input.

Simulation protocol. We simulate 200 trials in one block (Figure 3A). Every trial lasts for 10 simulated seconds (Figures 3B-E). In each trial, we apply different inputs to simulate different conditions as follows. First, we simulate the first to the 99th trial with rewarding CS and rewarding US – learning trials. The network can associate the rewarding CS to the rewarding US. The 100th trial is a "test" trial and the network receives rewarding CS and nonrewarding US. We then simulate the unexpected reward condition, i.e., nonrewarding CS and rewarding US. From the 101th trial to 199th trial, the network receives nonrewarding CS and nonrewarding US. The network associates the nonrewarding CS to the nonrewarding US. At the 200th trial, the network receives non-rewarding CS but rewarding US. See Figure 3A for a summary of the learning protocol.

We implement different inputs from the cortex to the VS and striosome based on the 4 conditions: reward CS, nonreward CS, reward US, and nonreward US. The rewarding/nonrewarding CS and US are shown in Figure 3 and their mathematical expressions are given in the Mathematics and Equations section. Note that the inputs from the cortex is always larger than zero value (firing rate activity cannot be negative in value).

The motivation for such an implementation is based on some observed evidences. First, neurons in the orbitofrontal cortex fire most strongly for cues that predict large

reward (with small penalty) and least strongly for cues that predict large-penalty (with small reward) relative to neutral conditions (small reward and small penalty) (Roesch and Olson, 2004; Morrison and Salzman, 2011). Second, cortical neurons, including the frontal cortex, are known to exhibit flexibility and mixed response properties, i.e. different cortical neurons could have different response to identical stimuli (Mante et al.,2013; Fusi et al.,2016). For instance, an identical tone could result in different responses from different cortical neurons which could in turn separately transmit information to the same neurons "downstream" e.g. in the midbrain. Third, the expectation values of cue signalling are stored in the cortex but not in the basal ganglia or LHb (Padoa-Chioppa and Assad,2006; Padoa-Chioppa and Conen,2017). The phasic activity of DA neurons can result in plasticity in the cortex and change the representation of cue signaling (Pascoli,et al, 2015). In fact, the activity profile in Figures 3D and E look similar to that of DA release or non-release (as measured e.g. in voltammetry (Phillips et al., 2003)). Also, the sustained or persistent activity in Figure 3B could represent (working) memory of the cue, a commonly observed phenomenon in the frontal cortical neurons (Miller et al.,1996; Padoa-Chioppa and Assad,2006; Padoa-Chioppa and Conen,2017), while the suppressed activity in Figure 3C can be thought of as some inhibitory effect with respect to the response in Figure 3B.

Results

Shift of phasic response from US to CS

Many experimental and theoretical studies have reported the shift of DA neurons response from US to CS (Ljungberg et al., 1992; Schultz, 1998; Pan et al., 2005). As discussed previously, the initial phase of learning, DA neurons are phasically activated from baseline upon the presentation of an unpredicted reward. An accompanied cue is associated to the rewarding outcome through a learning process. After learning, the phasic activity at reward outcome subsequently decreases to baseline, while a phasic activity now appears upon cue onset (Figure 1).

Our simulation can replicate this trend (Figure 4). When the network receives the rewarding CS and rewarding US (during the first 99 trials), DA neurons exhibit phasic activity upon the US in the first trial (Figure 4A). In the second and the subsequent trials, the peak appears upon the CS onset and the previous peak activity upon US onset disappears (Figures 4B and C).

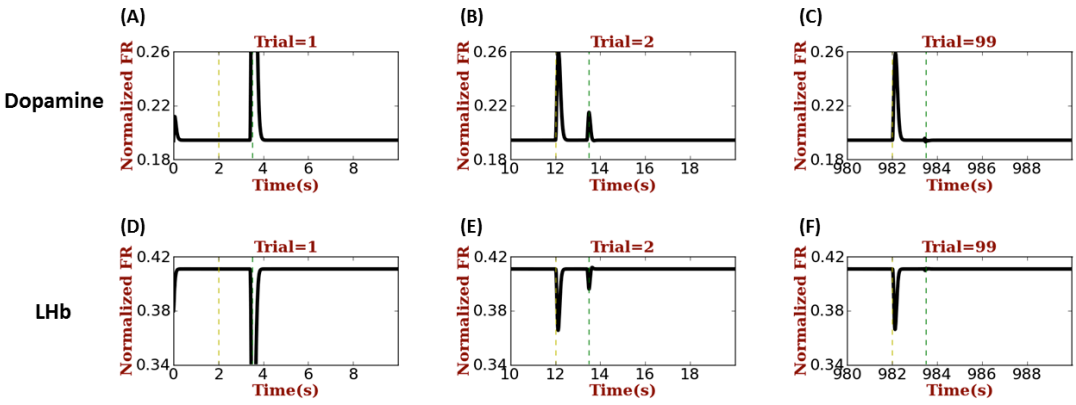

Figure 4. The shift of DA and LHb neurons' responses from US to CS. At the beginning of our simulation, the model circuit receives a reward CS and a reward US. FR: Neural firing rate activity. **(A)** Response of dopamine neurons in the first trial: DA neurons exhibit a phasic peak upon US and does not respond to CS in the first trial. **(B)** Response of DA neurons in the second trial: the activity of DA neurons shows a peak upon CS, and a peak

upon US. The response upon US is weaker than the response in the 1st trial. The responses of DA neurons in the 3rd to 98th trials are similar to (B), but the peak upon US gets weaker over trials. **(C)** Response of DA neurons in the 99th trial: the activity of DA neurons shows a peak upon CS, but baseline responding to US after learning. **(D)** Response of LHb neurons in the first trial: LHb neurons exhibit a phasic dip upon US and does not respond to CS. **(E)** Response of LHb neurons in the second trial: the activity of LHb neurons shows a dip upon CS, and a dip upon US. The response upon US is weaker than the response in the 1st trial. The responses of DA neurons in the 3rd to 98th trials are similar to (E), but the dip upon US get weaker trial by trial. **(F)** Response of LHb neurons in the 99th trial: the activity of LHb neurons shows a dip upon CS, but baseline responding to US after learning. (A) (B) (C) show the shift of DA neural response from US to CS after learning, while (D) (E) (F) show the shift of LHb neural response. The yellow dashed line indicates the time at which CS appears and the green dashed line indicates the time at which rewards are released or not.

The parallel pathways in our model can account for the shift in neural response from US to CS. At the beginning of the learning phase, CS-to-VS synaptic weights W_{iS} and CS input-to-striosomal synaptic weights Z_{ij} are very small or near zero. Thus, the activity of striosome is maintained at baseline level but the activity of VS has a peak upon US onset. The peak activity of VS then propagates to the LHb through the VS-VP-GPb-LHb pathway, which results in a dip of the LHb activity upon US. Meanwhile, a phasic input to DA neurons through the VS-VP-GPb-LHb-RMTg-VTA/SNc pathway and VS-PPTN-VTA/SNc pathway leads to a phasic activity of DA neurons upon reward US. The phasic activity of DA neurons upon reward US in turn enhances the positive reinforcement learning signal

N^+ (Eq.7) which leads to stronger synaptic strengths of afferent inputs to VS and striosome from the cortex: the increased synapse W_{iS} and Z_{ij} will enhance CS signal pathways from VS to DA via the PPTN (VS-PPTN-VTA/SNc) and VP (VS-VP-GPb-LHb-RMTg-VTA/SNc), the pathway from striosome to DA (striosome-VTA/SNc), and the pathway from striosome to DA via GPb (striosome-GPi-GPb-LHb-RMTg-VTA/SNc).

The striosome in the model has an adaptive timing spectrum, encoding the timing and the amount of reward associated with the CS (Fiala et al., 1996; Brown et al., 1999; Burke and Tobler, 2017) (see Eqs 10-14). Therefore, through the VS-PPTN-VTA/SNc pathway, rewarding CS can trigger phasic activity of DA neurons (Figures 4A-C) while nonrewarding CS can trigger a dip in activity (Figures 5C-D). The signal of rewarding US through the striosome inhibits DA neurons at the time when the rewarding US is expected to be present, but the excitation of reward US through the VS to VTA/SNc pathway via PPTN cancels the inhibition of the CS, leading to a baseline activity of DA neurons to reward US (Figure 4C, Figure 5A). On the contrary, nonrewarding US cannot trigger enough excitation to cancel the inhibition caused by CS in DA neurons, leading to a dip activity upon nonrewarding US onset (Figure 5B).

Experimental studies have shown that the phasic activity of LHb is opposite to that of DA neurons in terms of response to reward valence, but a similar shift in activity as

DA phasic activity. In our model, LHb neurons are inhibited and show a dip in their activity upon rewarding US onset (Figure 4D). The dip of LHb neural activity shifts from US to rewarding CS in the following and subsequent trials (Figures 4E-F). As mentioned previously, unexpected rewarding US can switch on the pathways: striosome-GPi-GPb-LHb and VS-VP-GPb-LHb. However, before they are switched on, the rewarding US will inhibit LHb neurons through the VS-VP-GPb-LHb pathway (Figure 4D). Once the striosome-LHb pathway and VS-LHb pathways are switched on, the reward CS will effectively inhibit LHb neurons through the VS-VP-GPb-LHb pathway, leading to a dip at the time of the rewarding CS. But the inhibition caused by the rewarding US will be canceled by excitation from the striosome-GPi-GPb-LHb pathway leading to a baseline activity of LHb neurons at the time of the rewarding US (Figure 4F).

Neural pathways underlying learned phasic activity of DA neurons

The phasic activity of DA neurons has been suggested to encode reward prediction error and play a pivotal role in reinforcement learning (Schultz et al. 1997; Morris et al. 2004; Bayer and Glimcher, 2005). DA neural activity in our model shows reward prediction error that is consistent with experimental observations (Figure 5F). For instance, after 99 trials of training, the network already can associate the rewarding CS to the rewarding US. The DA neurons show a phasic activity upon CS onset (at time 2 s in Figure 5A). But at the 100^{th} trial, we simulate the condition where the expected reward is omitted. DA neurons are excited right after CS onset (2 s) while inhibited at US presentation (3.6 s) (Figure 5B). The network now re-associates the

CS with the nonrewarding US after the training from the 101th to 199th trials. The activity of DA neurons then shows a dip at the time when nonrewarding CS is presented at 2 s and shows baseline activity when the nonrewarding US is presented at 3.6 s (Figure 5C). Finally, at the 200th trial, we present both the nonrewarding CS and rewarding US to simulate an unexpected reward condition. DA neurons are inhibited upon CS presentation (2 s) but excited at the time when rewarding US is presented once again (3.6 s) (Figure 5D). The overall activity profile of DA neurons is summarized in Figure 4E, which are consistent with experimental observations (Figure 5F).

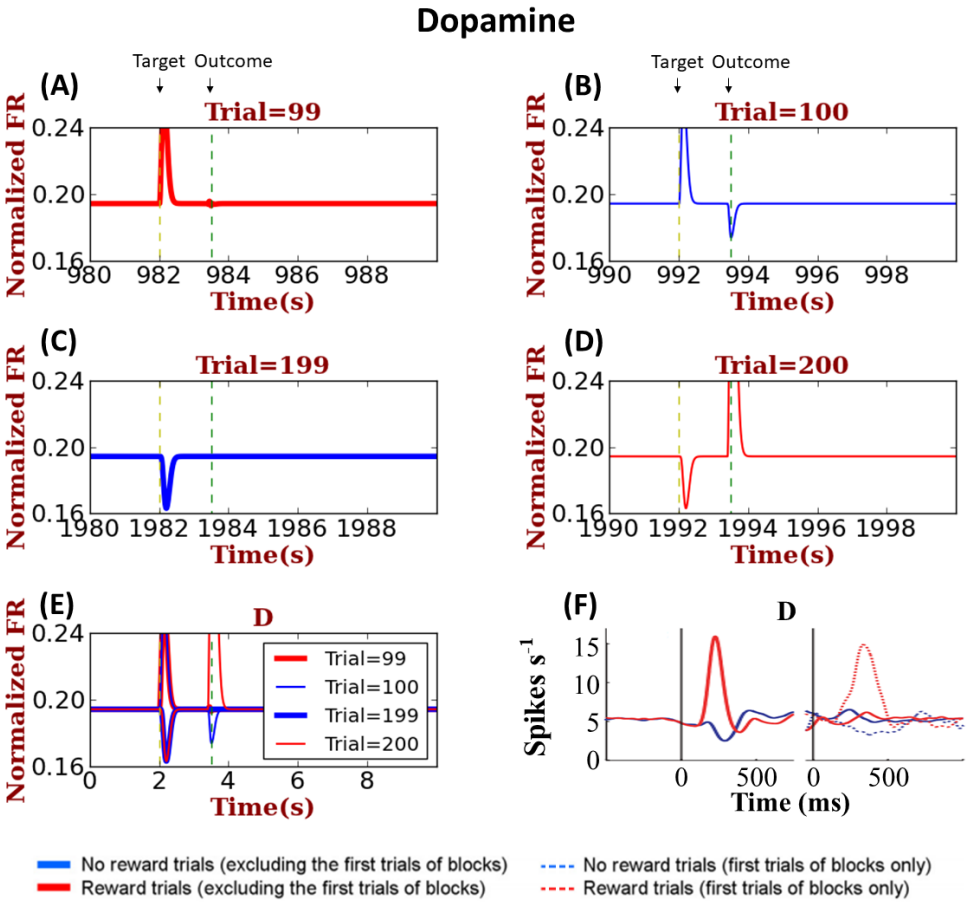

Figure 5. Acquired response of DA neurons. (A) The 99th trial: From the first to 99th trials, the model circuit receives a rewarding CS and a rewarding US. The result shows that after learning, DA neurons exhibit a phasic

peak upon rewarding CS and a baseline in response to reward outcome. **(B)** The 100^{th} trial: The model circuit receives rewarding CS and nonrewarding US. The result shows that DA neurons exhibit a phasic peak when rewarding CS appear and exhibit a phasic dip at the time the reward is expected. **(C)** The 199^{th} trial: From 101^{st} to 199^{th} trials, the model circuit receives nonrewarding CS and a nonrewarding US. The result shows that after learning, the DA neurons exhibit a phasic dip upon nonrewarding CS and a baseline when there is no reward released at this trial. **(D)** The 200^{th} trial: the model circuit receives nonrewarding CS and rewarding US. The result shows that DA neurons exhibit a phasic dip when nonreward CS appear and exhibit a phasic peak upon reward US. **(E)** The phasic activity of DA neurons under different situations. The thick red line indicates the activity of DA neurons at 99^{th} trial, the narrow blue line indicates the activity of DA neurons at 100^{th} trial, the thick blue line indicates the activity of DA neurons at 199^{th} trial, and the narrow red line indicates the activity of DA neurons at the 200^{th} trial. The yellow dashed line indicates the time at which CS appear and the green dashed line indicates the time at which rewards are released or not. **(F)** The physiological experimental result reprinted with permission from Matsumoto and Hikosaka (2007). Red lines indicate reward trials, and blue lines indicate no reward trials. Full lines indicate reward CS-to-reward US (red) and nonreward CS-to-nonreward US (blue), while dashed lines indicate reward CS-to-nonreward US (blue) and nonreward CS-to-reward US (red).

The above phasic responses of DA neural activity associated with the learned stimuli can be understood based on the two parallel pathways in the circuit: the VS-PPTN-VTA/SNc and striosome-VTA/SNc pathways. It should be noted that after the 1^{st} trial, the synaptic strengths W_{iS} and Z_{ij} are not zero, so VS responds to both rewarding CS and rewarding US. Then the DA neurons are excited by the rewarding CS through the VS-PPTN-VTA/SNc pathway. When rewarding US is presented, the

signal of rewarding CS triggers the activity of striosomal neurons and directly inhibits DA neurons. However, this inhibition is canceled out by the excitation from rewarding US through the VS-PPTN-VTA/SNc pathway. Thus the activity of DA neurons is effectively maintained at baseline (Figure 5A). By the 99th trial, the network has already associated the rewarding CS with rewarding US.

Now if the rewarding US is omitted (at the 100th trial), no excitation counterbalances the direct inhibition from striosome, leading to a dip of the activity of DA neurons (Figure 5B). This continues until the 199th trial. When the network is presented with a nonrewarding CS followed by nonrewarding US, the direct inhibitory pathway from striosome to DA neurons have been turned off, DA neurons show phasic activity upon nonrewarding CS onset while the activity of DA neurons is maintained at baseline at the time of nonrewarding US (Figure 5C). With a subsequently unexpected rewarding US in trial 200, DA neurons are now excited through the VS-PPTN-VTA/SNc pathway while the nonrewarding CS still causes a dip in the activity (Figure 5D).

Neural pathways underlying learned phasic activity of LHb neurons

Experimental studies have shown that phasic activity of LHb behaves in opposite way to that of DA neurons (Matsumoto and Hikosaka, 2007). Hence, it has been suggested that LHb neurons play key role in the coding of the aversive/negative signals (Meye et al., 2013; Song et al., 2017). Experiments have been carried out to investigate the activity of several brain nuclei, such as GPb (Hong and Hikosaka, 2008) and RMTg

(Hong et al., 2011), to explore the possible functional relationship with these brain regions.

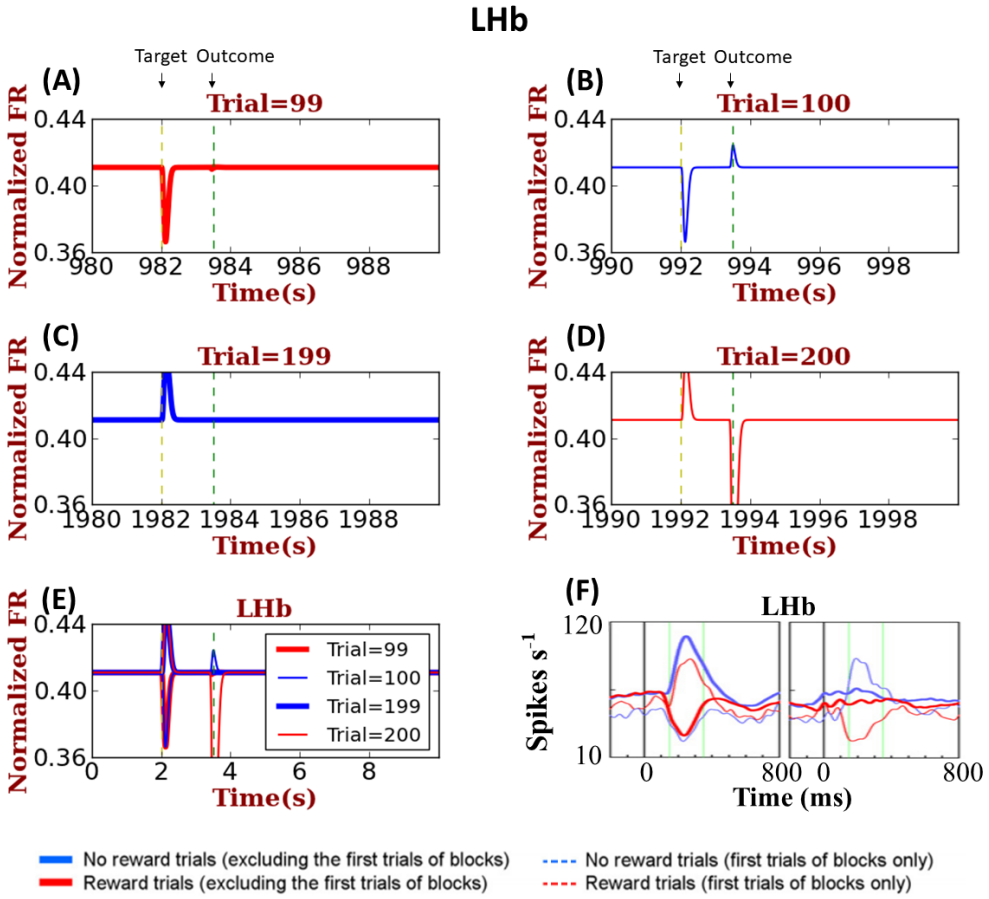

Figure 6. Acquired response of LHb neurons. (A) The 99th trial: From the 1st trial to the 99th trial, the model circuit receives rewarding CS and rewarding US. The result shows that after learning, LHb neurons exhibit a phasic dip upon rewarding CS and a baseline activity in response to rewarding outcome. **(B)** The 100th trial: The model circuit receives rewarding CS and nonrewarding US. The result shows that LHb neurons exhibit a phasic dip when rewarding CS appear and exhibit a phasic peak at the time when the reward should be released. **(C)** The 199th trial: From the 101th trial to the 199th trial, the model circuit receives nonrewarding CS and nonrewarding US. The result shows that after learning, LHb neurons exhibit a phasic peak upon nonrewarding CS and a baseline activity due to omission of reward at this trial. **(D)** The 200th trial: The model circuit receives nonrewarding CS and rewarding US. The result shows that LHb neurons exhibit a phasic peak when nonrewarding CS appear and

exhibit a phasic dip upon rewarding US. **(E)** The phasic activity of LHb neurons under different situations. The thick red line indicates the activity of LHb at the 99th trial, the narrow blue line indicates the activity of LHb at the 100th trial, the thick blue line indicates the activity of LHb at the 199th trial, and the narrow red line indicates the activities of LHb at the 200th trial. The yellow dashed line indicates the time at which CS appear and the green dashed line indicates the time at which rewards are released or not. **(F)** The physiological experimental results reprinted from Hong and Hikosaka (2008). Red lines indicate reward trials, and blue lines indicate no reward trials. Thick lines indicate reward CS-to-reward US (red) and nonreward CS-to-nonreward US (blue), while narrow lines indicate reward CS-to-nonreward US (blue) and nonreward CS-to-reward US (red).

Here, we simulate the activity of these nuclei and the results are consistent with the experimental observations. Our simulations show that the phasic responses of LHb neurons shift from US to CS. LHb neurons show a phasic dip when the unexpected rewarding US was presented in the first trial (Figure 4D). In the following trials, the dip shifts to the time when the rewarding CS presented (Figures 4E-F) and baseline activity is observed with rewarding CS (Figure 6A) and a small phasic activity upon nonrewarding US (Figure 6B). After the training of nonrewarding CS from the 101th to the 199th trials, LHb neurons show a phasic activity upon nonrewarding CS (2 s) while maintaining at baseline level at the time of the nonrewarding US (Figure 6C). At the 200th trial, LHb neurons show a peak activity with the nonrewarding CS but a big dip in activity given an unexpected rewarding US (Figure 6D). The overall activity profile of LHb neurons (Figure 6E) agrees with the experimental observations (Figure 6F).

The above mentioned learned phasic activity of LHb neurons can be explained with the two parallel pathways: striosome to LHb pathway via GPi and GPb and the VS to LHb pathway via VP and GPb. For instance, at the 99th trial, the synaptic strengths W_{iS} and Z_{ij} are not zero, which means that the network has already completely associated the rewarding CS with rewarding US. The rewarding CS can inhibit LHb neurons through the inhibitory striatum-VP-GPb-LHb pathway. When the rewarding US appears, the inhibition through the striatum-VP-GPb-LHb pathway will be canceled out by the excitation from the striosome-GPi-GPb-LHb pathway, resulting in a baseline level of LHb neural activity upon reward omission. At the 100th trial, LHb neurons show a dip in the presence of the rewarding CS. But the omission of reward implies that the excitation through striosome-GPb-LHb pathway cannot be canceled out, which leads to a small phasic activity of LHb neurons upon reward omission. At the same time, the synaptic strength Z_{ij} from cortex to striosome decreases to zero. When next the nonrewarding CS is paired with a nonrewarding US (from the 101th to 200th trial), LHb neurons show a phasic activity at the time of the nonrewarding CS onset because of the inhibition through the striatum-VP-GPb-LHb pathway. In the 200th trial, unexpected rewarding signal switches on the inhibitory pathway striosome-GPb-LHb, which leads to a dip in activity of the LHb neurons.

Learned phasic activity of GPb and RMTg

Experiments have shown that the GPb and RMTg neurons display phasic responses to CS and US. In our model, the interaction between striosome-GPi-GPb pathway and VS-VP-GPb pathway leads to the phasic activity of GPb neurons upon CS and US presentation. In particular, the GPb, LHb and RMTg are also connected with effectively excitatory synapses (Figure 2), and hence their phasic activities should be correlated with that of the LHb, with the same explanations of activity profiles as for the LHb (Figures 7 and 8). Moreover, the LHb-RMTg-VTA/SNc pathway only magnifies the phasic activity of DA neurons and does not qualitatively change the activity profile of DA neurons.

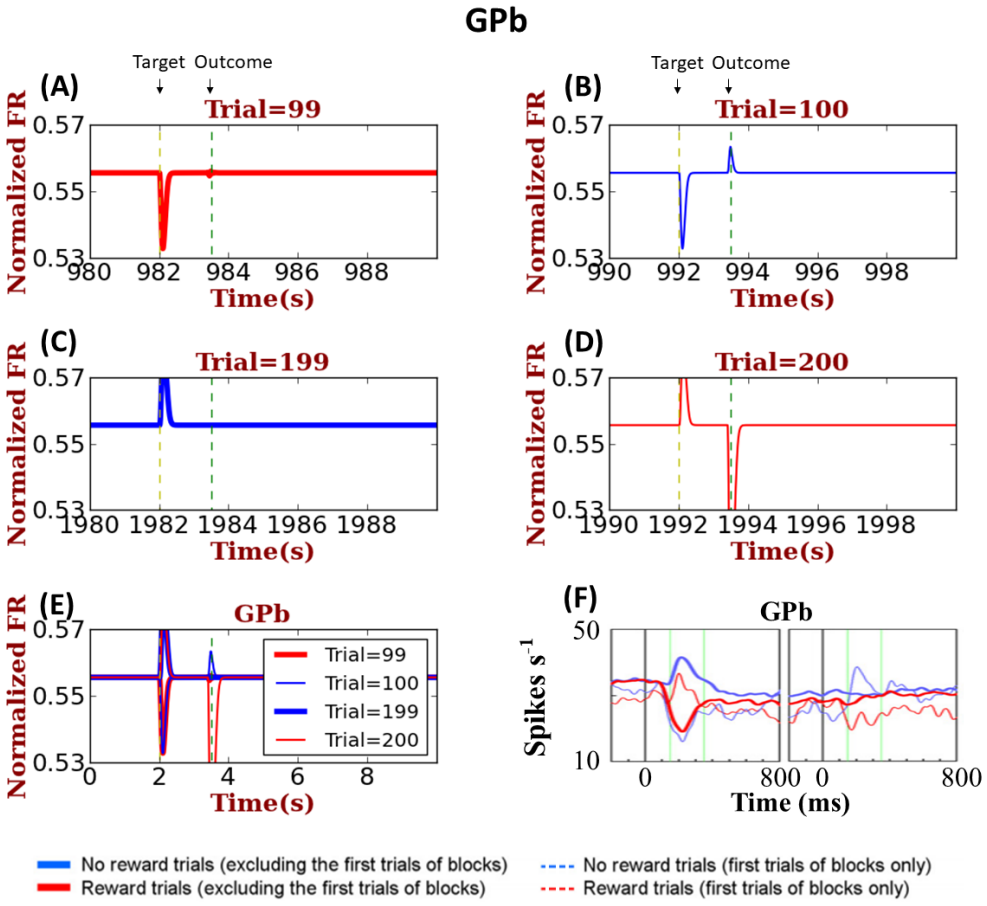

Figure 7. Acquired response of GPb neurons. (A) ~ (E): similar to Figure 6. **(F)** The physiological experimental result reprinted from Hong and Hikosaka (2008).

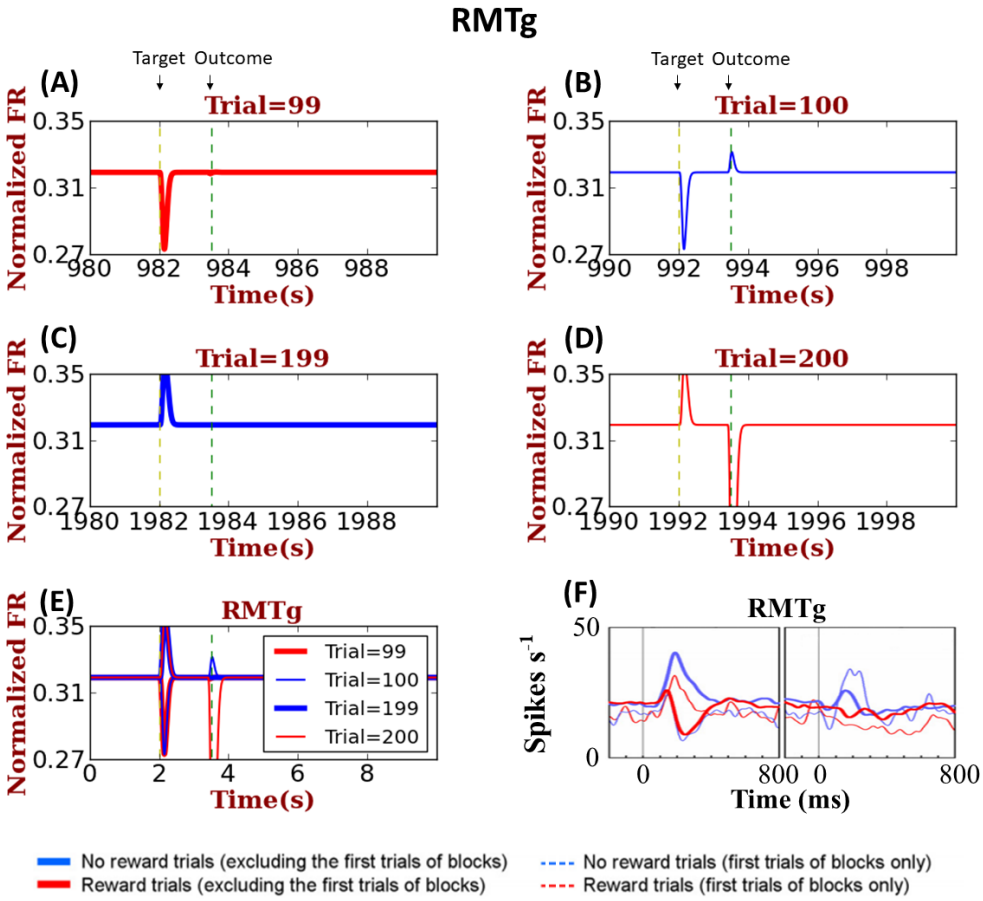

Figure 8. Acquired response of RMTg neurons. **(A) ~ (E)**: similar to Figure 6. **(F)** The physiological experimental result from Hong et al. (2011).

Robustness analysis of two parallel pathway model

Having shown the important role of the parallel circuit pathways in reproducing the phasic activities observed in experiments, we next further investigate the robustness of the phasic activities in our model with respect to connectivity strength variation.

Specifically, we increase or decrease all synaptic weights by 10% and monitor how the phasic activities change.

First, we found that the phasic activities of DA and LHb neurons did not change substantially when we increased or decreased the following synaptic weights by 10%: W_{SVP},W_{RS},W_{SP},W_{PD},W_{SOG},A_Z, and C_{WS}^{max} (data not shown). Second, weights of synapses on the pathway VP-GPb-LHb-RMTg-VTA/SNc was found to influence the tonic baseline activity of DA neurons, which we define as $\overline{D}$. Hence we change $\overline{D}$ while maintaining the phasic activity of DA and LHb neurons when we increase or decrease the weights of the synapses along this pathway (see Table 3). In Figures 9 and 10, we show the activity of DA neurons and LHb neuron given three different sets of synaptic weights from VP to GPb and corresponding baseline activities $\overline{D}$. We can see that DA and LHb neurons continue to demonstrate their characteristic phasic activity profiles. In brief, our neural circuit model is robust to the variation of synaptic weights.

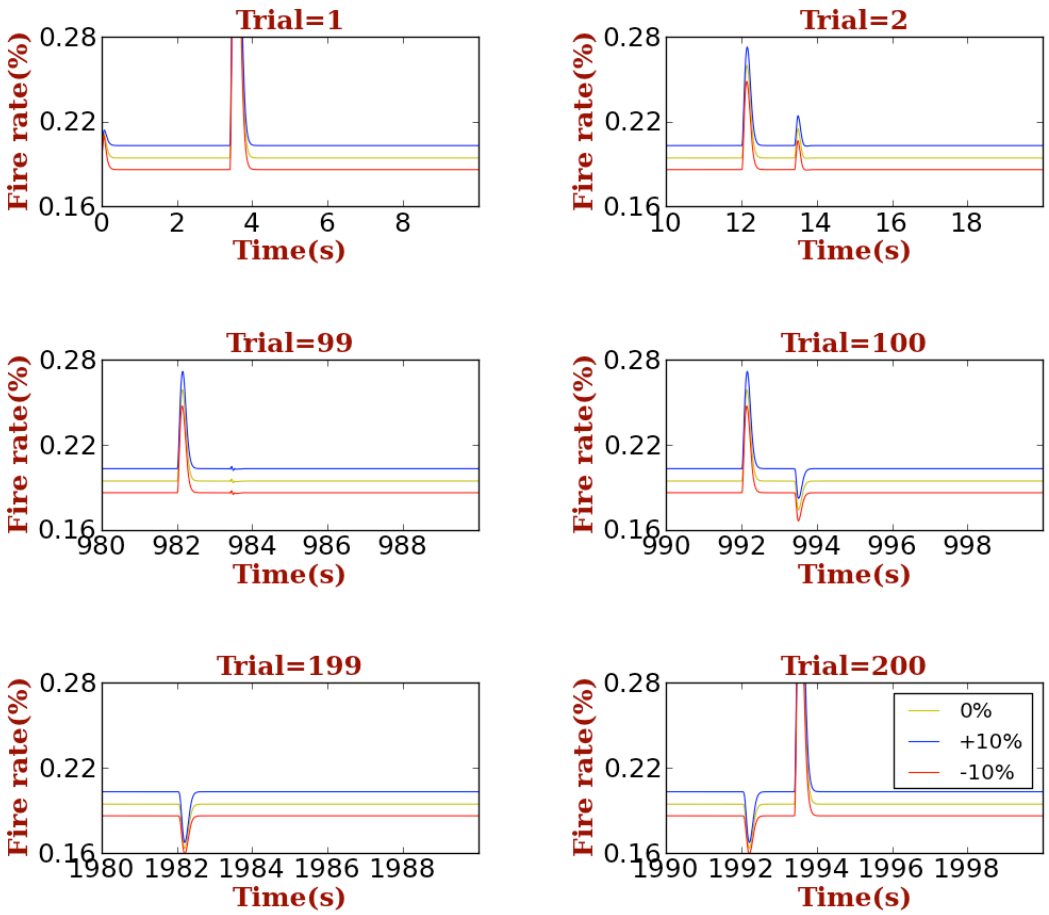

Fig 9. The phasic activity of DA neurons given different weight of synapse from VP to GPb. Yellow lines indicate the activity of DA neurons when W_{VPG} equals 1.00 and $\bar{D}$ equals 0.19431, blue lines indicate the activity when W_{VPG} equals 1.10 and $\bar{D}$ equals 0.20307, and red lines indicate the activity when W_{VPG} equals 0.90 and $\bar{D}$ equals 0.18608. **(A)** Trial 1: Phasic peak activity responds to unconditional reward. **(B)** Trial 2: The phasic activity shifts to the cue. **(C)** Trial 99: The phasic activity upon the cue and baseline activity upon the reward. **(D)** Trial 100: The dip activity upon reward omission. **(E)** Trial 199: The dip activity upon nonrewarding cue. **(F)** Trial 200: The peak activity upon unexpected reward.

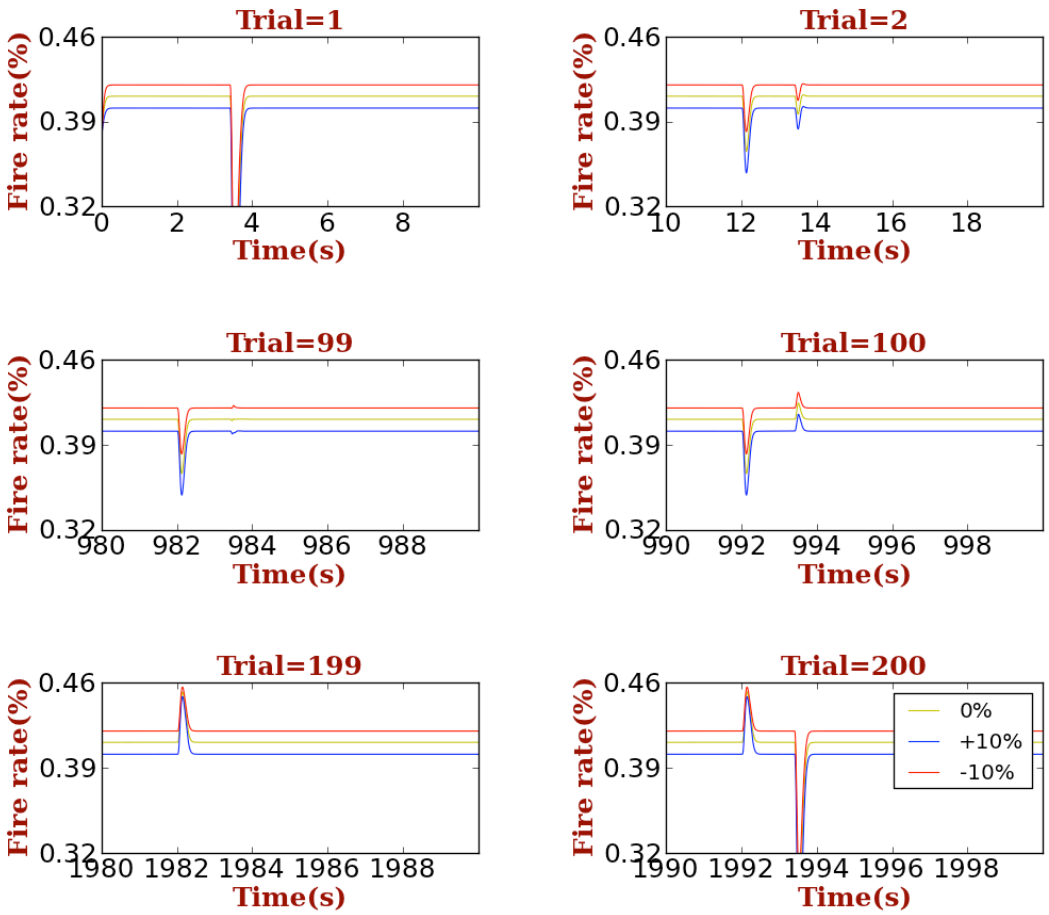

Figure 10. The phasic activity of LHb neurons given different weight of synapse from VP to GPb. Yellow lines indicate the activity of LHb neurons when W_{VPG} equals 1.00 and $\bar{D}$ equals 0.19431, blue lines indicate the activity when W_{VPG} equals 1.10 and $\bar{D}$ equals 0.20307, and red lines indicate the activity when W_{VPG} equals 0.90 and $\bar{D}$ equals 0.18608. **(A)** Trial 1: Phasic dip activity responds to unconditional reward. **(B)** Trial 2: The phasic activity shifts to the cue. **(C)** Trial 99: The phasic activity upon the cue and baseline activity upon the reward. **(D)** Trial 100: The peak activity upon reward omission. **(E)** Trial 199: The peak activity upon nonrewarding cue. **(F)** Trial 200: The dip activity upon unexpected reward.

Discussion

We extended a previous neural circuit model (Brown et al., 1999) by incorporating the nuclei GPb, LHb and RMTg, and the model could account for various experimental data from separate works. Specifically, the model could exhibit the shift of DA and LHb neural responses from US to CS presentation times. Our simulations also replicated the phasic activity of DA, LHb, GPb and RMTg neurons observed in experiments. The DA (LHb) neurons exhibited a phasic peak (dip) upon reward CS, and maintenance of baseline activity in response to a rewarding outcome but a phasic dip (peak) if the reward is omitted. By contrast, the DA (LHb) neurons exhibited a phasic dip (peak) in response to a nonrewarding CS or punishment CS, and maintenance of baseline activity in response to the nonrewarding US, but a phasic peak (dip) if a reward occurs or the aversive US is omitted. The acquired responses of GPb and RMTg neurons are similar to that of LHb neurons. These acquired responses are consistent with experimental data (W. Schultz et al., 1997; Matsumoto and Hikosaka, 2007; Hong and Hikosaka, 2008; Hong et al., 2011) and behavioral experiments (Danna et al., 2013).

Our model provides insights to the neural circuit mechanism of DA and LHb phasic activity. In particular, parallel excitatory and inhibitory pathways underlie the learned responses: striatum-to-PPTN-to-VTA/SNc pathway excites DA, while striosome--VTA/SNc pathway inhibits DA; striatum-to-VP-to-GPb-to-LHb pathway

inhibits LHb, while striosome-to-GPb-to-LHb pathway excites LHb; LHb-to-RMTg-to-VTA/SNc pathway magnifies the phasic activity of DA. Under different task conditions, we apply different CS input and US input to the model. The model has a feedback loop in which DA can modulate the cortico-striatal synapses and the cortico-striosome synapses. This will in turn affect the DA responses, closing the loop. After learning, the weights of these synapses stabilize and remain unchanged. This led to the emergent phasic activity profiles of the nuclei in the circuit – with the parallel pathways balancing out one another. In addition, we found striosome to be a key brain nucleus which remembers the timing of previous rewards and encodes the predicted rewards. In fact, there is a recent experimental works (Takahashi et al., 2016) that supports our model prediction.

In our model, we predict neurons in the striosome to encode expected reward, but there are alternative theories. For example, Cohen et al. (2012) found that were three types of VTA neurons and VTA GABAergic neurons may signal expected reward, which could be a key variable for dopaminergic neurons to calculate reward prediction error. Recent works (Stauffer, 2015; Yoo et al., 2016; Morales and Margolis, 2017) highlight the importance of VTA GABAergic neurons. Averbeck and Costa (2017) proposed that the amygdala can learn and represent expected values like the striatum, and they predicted that the amygdala may play a central role in reinforcement learning and the ventral striatum may play less of a primary role. Wagner et al. (2017) suggested that the cerebellar granule cells may encode the

expectation of reward. Luo et al. (2015), Li et al. (2016) and Hayashi et al.(2015) found that serotonin neurons in the dorsal raphe nucleus can encode reward signals. Some physiological and theoretical works (Tan and Bullock, 2008; Humphries and Prescott, 2010; Keiflin and Janak, 2015; Hikida et al., 2016) focus on D1 and D2 receptors in the ventral striatum and suggested that they play important role in computing reward prediction error. Future neural circuit modeling effort would need to incorporate such findings.

To obtain the results consistent with experiments, we have adopted several assumptions. First, we assumed that the striatal neurons excite the PPTN and ventral pallidum. Striatal neurons are usually identified as GABAergic and inhibitory, but they may excite downstream neurons through disinhibitory effect or substance P released by striatal neurons (Napier et al., 1995; Blomeley et al. 2009). In fact, it has been demonstrated that substance P mediates the excitatory interaction between striatal neurons to VP neurons (Napier et al., 1995) and striatal projection neurons (Blomeley et al. 2009). Second, we hypothesized that the striosome projects to the GPi which in turn projects to the GPb. Although we have no direct evidence, Hong and Hikosaka (2013) have observed that typical GPe and GPi neurons are first inhibited by striatal stimulation and GPb neurons are often (but not always) excited by striatal stimulation. They proposed that inputs to GPb were mediated through inhibitory axon collaterals within the striatum (Tremblay and Filion,1989) or GPe (Sadek et al., 2007).

While developing the model, we have tried to add minimal features to the previous Brown et al. (1999) model. Hence, it is worthy of note that we have ignored several factors to simplify the model. Specifically, we ignored the connections between some brain nuclei, such as the cortex-to-GPb (Hong and Hikosaka, 2008), VP-to-RMTg (Hong et al., 2011), LHb-to-LHb, and cortex-to-LHb (Meye et al., 2013), and DA-to-striatum (Parker et al., 2016) pathways. We also did not consider the direct LHb-to-VTA (Poller et al., 2013) and VTA-to-LHb (Stamatakis et al., 2013) connections in our simulation, but we mimicked the overall inhibition of LHb on VTA. We have also ignored the different types of activity of many brain nuclei. For instance, studies have suggested three types of GPb neurons: reward-postive type, reward-negative type and direction selective type (Hong and Hikosaka, 2008). Our model only considers the reward-negative type since the majority of the reward-negative type is in the GPb and may be key to

Despite the assumptions in the model, our neural circuit model can still implement the computation for reward based phasic signaling and reinforcement learning, as observed in a variety of experiments. The phasic activities in multiple brain regions represent prediction error signals, which not only associates the cue to outcome but also memorizes the specific time interval between the two. This requires the neural system to hold the information predicted by the cue, compare the information with the outcome, and report the result of the comparison. In our model, the time spectrum of

the striosome and the parallel excitatory and inhibitory pathways provided the platform for such computation. The peak activity of DA and LHb neurons function in complementary roles–encoding reward and nonreward/punishment information separately–and alleviating any flooring (limiting) effect of the dip in activity of either neuron types. Our novel neural circuit model with parallel pathways provides an instantiation instance of such complex neural computation.

References

Averbeck BB, Costa VD (2017) Motivational neural circuits underlying reinforcement learning. Nat Neurosci 20:505-512.

Bayer HM, Glimcher PW (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 47:129-141.

Blomeley CP, Kehoe LA, Bracci E (2009) Substance P mediates excitatory interactions between striatal projection neurons. J Neurosci 29:4953-4963.

Brinschwitz K, Dittgen A, Madai VI, Lommel R, Geisler R, Veh RW(2010) Glutamatergic axons from the lateral habenula mainly terminate on GABAergic neurons of the ventral midbrain. Neuroscience, 168(2):463-476.

Bromberg-Martin ES, Matsumoto M, Hikosaka O (2010) Dopamine in motivational control: rewarding, aversive, and alerting. Neuron 68:815-834.

Brown J, Bullock D, Grossberg S (1999) How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. J Neurosci 19:10502-10511.

Burke CJ and Tobler PN (2017) Time, not size, matters for striatal reward predictions to dopamine. Neuron 91(1):8-11.

Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N (2012) Neuron-type-specific signals for reward and punishment in the ventral tegmental area. Nature 482:85-109.

Danna CL, Shepard PD, Elmer GI (2013) The habenula governs the attribution of incentive salience to reward predictive cues. Front Hum Neurosci 7.

Fiala JC, Grossberg S, Bullock D (1996) Metabotropic glutamate receptor activation in cerebellar Purkinje cells as substrate for adaptive timing of the classically conditioned eye-blink response. J Neurosci 16:3760-3774.

Fiorillo CD, Tobler PN, Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. Science 299:1898-1902.

Fusi S, Miller EK, Rigotti M(2016) Why neurons mix: high dimensionality for higher cognition. Current Opinion in Neurobiology. 37: 66-74

Glimcher PW (2011) Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. Proc. Natl Acad Sci USA 108:17568.

Hayashi K, Nakao K, and Nakamura K(2015) Appetitive and aversive information coding in the primate Dorsal Raphé nucleus. J Neurosci. 35 (15) 6195-6208.

Hikida T, Morita M, Macpherson T (2016) Neural mechanisms of the nucleus accumbens circuit in reward and aversive learning. Neurosci Res 108:1-5.

Hikosaka O (2010) The habenula: from stress evasion to value-based decision-making. Nat Rev Neurosci 11:503-513.

Hong S, Hikosaka O (2008) The globus pallidus sends reward-related signals to the lateral habenula. Neuron 60:720-729.

Hong S, Hikosaka O (2013) Diverse sources of reward value signals in the basal ganglia nuclei transmitted to the lateral habenula in the monkey. Front Hum Neurosci 7.

Hong S, Jhou TC, Smith M, Saleem KS, Hikosaka O (2011) Negative reward signals from the lateral habenula to dopamine neurons are mediated by rostromedial tegmental nucleus in primates. J Neurosci 31:11457-11471.

Humphries MD, Prescott TJ (2010) The ventral basal ganglia, a selection mechanism at the crossroads of space, strategy, and reward. Prog Neurobiol 90:385-417.

Jhou TC, Fields HL, Baxter MG, Saper CB, Holland PC (2009) The rostromedial tegmental nucleus (RMTg), a GABAergic afferent to midbrain dopamine neurons, encodes aversive stimuli and inhibits motor responses. Neuron 61:786-800.

Kalen P, Pritzel M, Nieoullon A, Wiklund L (1986) Further evidence for excitatory amino acid transmission in the lateral habenular projection to the rostral raphe nuclei: lesion-induced decrease of high affinity glutamate uptake. Neurosci Lett 68:35-40.

Keiflin R, Janak PH (2015) Dopamine Prediction Errors in Reward Learning and Addiction: From Theory to Neural Circuitry. Neuron 88:247-263.

Kuhnen CM, Knutson B (2005) The neural basis of financial risk taking. Neuron 47:763-770.

Li Y, Zhong W, Wang D, Feng Q, Liu Z, Zhou J, Jia C, Hu F, Zeng J, Guo Q, Fu L, Luo M (2016) Serotonin neurons in the dorsal raphe nucleus encode reward signals. Nat Commun 7.

Ljungberg T, Apicella P, Schultz W (1992) Responses of monkey dopamine neurons during learning of behavioral reactions. J Neurophysiol 67:145-163.

Luo M, Zhou J, Liu Z (2015) Reward processing by the dorsal raphe nucleus: 5-HT and beyond. Learn Memory 22:452-460.

Mante V, Sussillo D,Shenoy KV, Newsome WT(2103) Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature 503:78-84.

Matsumoto M, Hikosaka O (2007) Lateral habenula as a source of negative reward signals in dopamine neurons. Nature 447:1111.

Matsumoto M, Hikosaka O (2009) Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature 459:834-837.

McCoy AN, Platt ML (2005) Risk-sensitive neurons in macaque posterior cingulate cortex. Nat Neurosci 8:1220-1227.

Meye FJ, Lecca S, Valentinova K, Mameli M (2013) Synaptic and cellular profile of neurons in the lateral habenula. Front Hum Neurosci 7.

E.K. Miller EK, Erickson CA, Desimone R(1996) Neural mechanisms of visual working memory in prefrontal cortex of the macaque. J. Neurosci. 16:5154-5167

Neural mechanisms of visual working memory in prefrontal cortex of the macaque

J. Neurosci., 16 (1996), pp. 5154-5167

Monosov IE, Hikosaka O (2013) Selective and graded coding of reward uncertainty by neurons in the primate anterodorsal septal region. Nat Neurosci 16:756.

Morales M, Margolis EB (2017) Ventral tegmental area: cellular heterogeneity, connectivity and behaviour. Nat Rev Neurosci 18:73-85.

Morris G, Arkadir D, Nevet A, Vaadia E, Bergman H (2004) Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. Neuron 43:133-143.

Morrison SE and Salzman CD(2011) Representations of appetitive and aversive information in the primate orbitofrontal cortex. Ann. N Y Acad. Sci.1239, 59–70.

Napier Tc, Mitrovic I, Churchill L, Klitenick Ma, Lu Xy, Kalivas Pw (1995) Substance-P in the ventral pallidum - projection from the ventral striatum, and electrophysiological and behavioral consequences of pallidal substance-P. Neuroscience 69:59-70.

Niv Y, Duff MO, Dayan P (2005) Dopamine, uncertainty and TD learning. Behavioral and brain functions: BBF 1.

Padoa-Schioppa C, Assad JA (2006) Neurons in the orbitofrontal cortex encode economic value. Nature 441, 223–226.

Padoa-Schioppa C,Conen KE(2017) Orbitofrontal Cortex: A Neural Circuit for Economic Decisions. Neuron 96(4): 736-754.

Pan WX, Schmidt R, Wickens JR, Hyland BI (2005) Dopamine cells respond to predicted events during classical conditioning: Evidence for eligibility traces in the reward-learning network. J Neurosci 25:6235-6242.

Parker NF, Cameron CM, Taliaferro JP, Lee J, Choi JY, Davidson TJ, Daw ND, Witten IB (2016) Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. Nat Neurosci 19:845.

Pascoli V, Terrier J,Hiver A, Lüsche C(2015) Sufficiency of mesolimbic dopamine neuron stimulation for the progression to addiction. Neuron 88:1054–1066

Phillips P.E.M., Robinson D.L., Stuber G.D., Carelli R.M., Wightman R.M. (2003) Real-time measurements of phasic changes in extracellular dopamine concentration in freely moving rats by fast-scan cyclic voltammetry. In: Wang J.Q. (eds) Drugs of Abuse. Methods In Molecular Medicine™, vol 79. Humana Press

Poller WC, Madai VI, Bernard R, Laube G, Veh RW (2013) A glutamatergic projection from the lateral hypothalamus targets VTA-projecting neurons in the lateral habenula of the rat. Brain Res 1507:45-60.

Press WH, Flannery BP, Teukolsky SA, Vetterling WT(2007) "Section 17.1 Runge-Kutta Method", Numerical Recipes: The Art of Scientific Computing (3rd ed.), Cambridge University Press.

Rajakumar N, Elisevich K, and Flumerfelt BA(1993) Compartmental origin of the striato-entopeduncular projection in the rat. J. Comp. Neurol. 331, 286–296.

Roesch MR and Olson CR(2004).Neuronal activity related to reward value and motivation in primate frontal cortex. Science 304, 307–310.

Sadek AR, Magill PJ, Bolam JP (2007) A single-cell analysis of intrinsic connectivity in the rat globus pallidus. J Neurosci 27:6352-6362.

Schultz W (1998) Predictive reward signal of dopamine neurons. J Neurophysiol 80:1-27.

Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593-1599.

Shabel SJ, Proulx CD, Trias A, Murphy RT, Malinow R (2012) Input to the lateral habenula from the basal ganglia is excitatory, aversive, and suppressed by Serotonin. Neuron 74:475-481.

Shabel SJ, Proulx CD, Piriz J, Malinow R (2014) Mood regulation. GABA/glutamate co-release controls habenula output and is modified by antidepressant treatment. Science 345:1494–1498

Song M, Jo YS, Lee Y, Choi J (2017) Lesions of the lateral habenula facilitate active avoidance learning and threat extinction. Behav Brain Res 318:12-17.

Stamatakis AM, Jennings JH, Ung RL, Blair GA, Weinberg RJ, Neve RL, Boyce F, Mattis J, Ramakrishnan C, Deisseroth K, Stuber GD (2013) A unique population of ventral tegmental area neurons inhibits the lateral habenula to promote reward. Neuron 80:1039-1053.

Stauffer WR (2015) Systems neuroscience: shaping the reward prediction error signal. Curr Biol 25:R1081-R1084.

Sutton RS (1988) Learning to predict by the methods of temporal differences. Machine Learning 3:9-44.

Sutton RS, Barto AG (1981). Toward a modern theory of adaptive networks: Expectation and prediction. Psychological review, 88(2), 135-170.

Takahashi YK, Langdon AJ, Niv Y, Schoenbaum G (2016) Temporal specificity of reward prediction errors signaled by putative dopamine neurons in rat VTA depends on ventral striatum. Neuron 91:182-193.

Tan CO, Bullock D (2008) A local circuit model of learned striatal and dopamine cell responses under probabilistic schedules of reward. J Neurosci 28(40), 10062-10074.

Tan CO, Bullock D (2008) A dopamine-acetylcholine cascade: Simulating learned and lesion-induced behavior of striatal cholinergic interneurons. J Neurophysiol 100:2409-2421.

Tremblay L, Filion M (1989) Responses of pallidal neurons to striatal stimulation in intact waking monkeys. Brain Res 498:1-16.

van Duuren, E., van der Plasse, G., Lankelma, J., Joosten, R. N., Feenstra, M. G., and Pennartz, C. M. (2009). Single-cell and population coding of expected reward probability in the orbitofrontal cortex of the rat. J Neurosci 29(28), 8965-8976.

Wagner MJ, Kim TH, Savall J, Schnitzer MJ, Luo L (2017) Cerebellar granule cells encode the expectation of reward. Nature 544:96.

Yoo JH, Zell V, Gutierrez-Reed N, Wu J, Ressler R, Shenasa MA, Johnson AB, Fife KH, Faget L, Hnasko TS (2016) Ventral tegmental area glutamate neurons co-release GABA and promote positive reinforcement. Nat Commun 7.

Mathematics and Equations

This section lists the mathematical equations of the model (Figure 2). We give the model circuit different inputs to simulate different conditions. We use differential equations to simulate the firing rates (or the activity levels) of the neurons in different brain areas. The model variables are summarized in Table 1, and the fixed parameters are summarized in Table 2. Table 3 summarizes the synaptic strength changed to test the robustness of the model.

Table 1. Model variables

S	**The activation level of ventral striatal neurons**
I_i	The ith CS input signal
I_R	The US input signal
W_{iS}	CS-to-VS synaptic weights
G_{WS}	Calcium signal
N^+	Above-baseline dopamine burst signal
N^-	Below-baseline dopamine dip signal
x_{ij}	Striosomal metabotropic response
r_j	Striosomal activity buildup rate parameter
$[G_{ij}Y_{ij}]^+$	Striosomal calcium concentration
Z_{ij}	CS input-to-striosomal synaptic weights
$P_{pre-excite}$	The level of substance P exciting PPTN

$P_{pre-inhibit}$	The level of GABA inhibiting PPTN
$VP_{pre-excite}$	The level of substance P exciting VP
$VP_{pre-inhibit}$	The level of GABA inhibiting VP
P	The activation level of PPTN neurons
VP	The activation level of VP neurons
$input_{pre\_P}$	The net effect of substance P and GABA into PPTN
$input_{pre\_VP}$	The net effect of substance P and GABA into VP
GPb	The activation level of GPb neurons
LHb	The activation level of LHb neurons
$RMTg$	The activation level of RMTg neurons
D	The activation level of DA neurons

Table 2. Model parameters

Symbol	Description	Value
$background_{IC}$	Baseline of CS input	0.30
$background_{IR}$	Baseline of US input	0.20
τ	exponentially decaying time constant of CS/US input	20.0
τ_S	The time constant of VS neurons	36.0
τ_{WS}	The time constant of the change of weight W_{iS}	6
α_{WS}	CS-to-VS weight learning rate	13.0
$C_{WS}^{\max}$	Maximum CS-to-VSsynaptic weight	4.00
β_{WS}	CS-to-VSweight decay rate	13.00

$\bar{D}$	The baseline activation level of DA neurons	0.194
Γ_D	Phasic dopamine signal threshold	0.001
α_r	Striosomal spectrum spacing	16.5
β_r	Striosomal spectrum offset	30.9
α_G	Calcium activation rate	3.00
B_G	Calcium concentration maximum	5.00
Γ_G	Calcium spike threshold	0.37
β_G	Calcium passive decay rate	12.00
α_Y	Calcium recovery rate	0.108
β_Y	Activity-dependent calcium inactivation rate	48.0
Γ_Y	Calcium inactivation threshold	0.18
α_Z	Striosomal learning rate	500.00
Γ_S	Striosomal output threshold	0.27
A_Z	Maximum CS input-to-striosomal synaptic weight	20.0
B_Z	CS input-to-striosomal synaptic weight decay rate	40.0
τ_{P1}	The time constant of the change of $P_{pre-excite}$	36.00
τ_{P2}	The time constant of the change of $P_{pre-inhibit}$	6.00
W_{SP}	VS-to-pre-PPTN synaptic weight	1.00
τ_{VP1}	The time constant of the change of $VP_{pre-excite}$	36.00
τ_{VP2}	The time constant of the change of $VP_{pre-excite}$	6.00
W_{SVP}	VS-to-pre-VP synaptic weight	1.00
$background_P$	The background input to the PPTN	0.10

W_P	VS-to-PPTN input weight	3.00
τ_P	PPTN neurons response time constant	36.00
$background_{VP}$	The background input to the VP	0.10
W_{VP}	VS-to-VP input weight	3.00
τ_{VP}	VP neurons response time constant	36.00
Γ_{P12}	The difference signal threshold of the excitatory and inhibitory effects previous to PPTN	0.006
Γ_{VP12}	The difference signal threshold of the excitatory and inhibitory effects previous to VP	0.006
τ_{GPb}	GPb neuron response time constant	36.00
$background_{GPb}$	The background input to the GPb	0.60
W_{VPG}	VP-to-GPb synaptic weight	1.00
W_{SOG}	Striosome-GPb synaptic weight	0.35
τ_{LHb}	LHb neuron response time constant	36.00
$background_{LHb}$	The background input to the LHb	0.10
W_{GL}	GPb-to-LHb synaptic weight	5.00
Γ_{GPb}	GPb output signal threshold	0.45
τ_{RMTg}	RMTg neuron response time constant	36.00
$background_{RMTg}$	The background input to the RMTg	0.10
W_{LR}	LHb-to-RMTg synaptic weight	2.00
Γ_{LHb}	LHb output signal threshold	0.25
τ_D	DA neuron response time constant	36.00

$background_D$	The background input to the D	0.40
W_{RD}	RMTg-to-VTA/SNc synaptic weight	0.80
W_{PD}	PPTN-to VTA/SNc synaptic weight	1.00
Γ_P	PPTN output signal threshold	0.10
h_D	Maximum hyperpolarization of DA neurons	0.10

Table 3 Changes of synaptic weights to test the robustness of the phasic activity

Synaptic weight	**+10%**	**-10%**
$\boldsymbol{W_{VPG}}$	0.20307（4.508%）	0.18608（-4.235%）
$\boldsymbol{W_{GL}}$	0.17691（-8.955%）	0.21327（9.758%）
$\boldsymbol{W_{LR}}$	0.18006（-7.334%）	0.20875（7.431%）
$\boldsymbol{W_{RD}}$	0.16571（-14.719%）	0.22102（13.746%）

The mathematical expressions are below:

(i) Different inputs in each trial (Figure 2)

The cortex, especially, the orbitofrontal cortex(OFC) encodes the expectation future outcome and their response reflect the value conveyed by the combination of reward and punishment of the cue(Padoa-Chioppa and Assad,2006; Padoa-Chioppa and Conen,2017). Furthermore, OFC neurons fired most strongly for cues that predict large reward or small penalty and least strongly for cues that predict large penalty or small reward relative to neutral conditions (Roesch and Olson,2004; Morrison and

Salzman, 2011). Therefore, we set a larger value for rewarding cue and smaller but positive value for nonrewarding cue as follows.

Reward CS input:

$$I_{C-reward} = \begin{cases} background_{IC} & 0 <= t <= 2 \\ background_{IC} + 0.60 & 2 < t <= 3.60 \\ background_{IC} + 0.60e^{-\frac{1}{\tau}(t-3.60)} & t > 3.60 \end{cases} \quad (1)$$

We set $background_{IC} = 0.30$ and $\tau = 20$.

When the network receives a reward CS, the inputs from cortex increase abruptly and last until the time the expected reward should be given. Then, the inputs decay exponentially to baseline activity level.

Nonreward CS input:

$$I_{C-nonreward} = \begin{cases} background_{IC} & 0 <= t <= 2 \\ background_{IC} - 0.20 & 2 < t <= 3.60 \\ background_{IC} - 0.20e^{-\frac{1}{\tau}(t-3.60)} & t > 3.60 \end{cases} \quad (2)$$

Reward US input:

$$I_{R-reward} = \begin{cases} background_{IR} & 0 <= t <= 3.40 \\ background_{IR} + 0.80 & 3.40 < t <= 3.60 \\ background_{IR} + 0.80e^{-\frac{1}{\tau}(t-3.60)} & t > 3.60 \end{cases} \quad (3)$$

We set $background_{IR} = 0.20$.

When the network receives a reward US, the inputs from lateral hypothalamus increase abruptly and last for a very short duration. Then, the inputs decay exponentially to baseline activity level.

Nonreward US input:

$$I_{R-nonreward} = background_{IR} \tag{4}$$

If the network doesn't get reward, or get nonreward (aversion or punishment), we assume the inputs in this trial do not change, and the inputs remain at baseline level.

(ii) Differential equations

First, the changes of activation level of ventral striatal cells (S) are governed by (Brown et al., 1999):

$$\frac{1}{\tau_S}\frac{d}{dt}S = -S + (1-S)[\sum_i I_i W_{iS} + I_R W_{RS}] \tag{5}$$

The activity level of striatal cells changes in the wake of its passive decay and excitation from CS inputs and US inputs. The weight W_{RS} is fixed while the weight W_{iS} can be changed.

The weight W_{iS} is governed by (Tan and Bullock, 2008):

$$\frac{1}{\tau_{WS}}\frac{d}{dt}W_{iS} = G_{WS}S[\alpha_{WS}N^+ I_i(C_{WS}^{\max} - W_{iS}) - \beta_{WS}N^- W_{iS}] \tag{6}$$

The synaptic weight changes are induced by phasic dopamine burst or dip signal, N^+ and N^- (defined below in Equations 7 and 8). Learning is gated by delayed release

of a second messenger and calcium signal G_{WS} is governed by Equations 9 and 11 (below) at a rate r=12.5.

The positive reinforcement-learning signal N^+ derives from excitatory phasic fluctuations of the dopamine signal above the baseline:

$$N^+ = [D - \bar{D} - \Gamma_D]^+ \tag{7}$$

The complementary negative reinforcement-learning signal N^- derives from inhibitory phasic fluctuations of the dopamine signal below baseline:

$$N^- = [\bar{D} - D - \Gamma_N]^+ \tag{8}$$

Second, strisomes play an important role in the phasic activities of DA neurons and LHb neurons because of its timing spectrum mechanism: a spectrum of striosomal MSPN second messenger activities x_{ij} respond to the ith input at rates r_j :

$$\frac{d}{dt}x_{ij} = r_j[-x_{ij} + (1 - x_{ij})I_i] \tag{9}$$

where the second messenger buildup rates are given by

$$r_j = \frac{\alpha_r}{\beta_r + j} \tag{10}$$

The activities x_{ij} induce intracellular calcium dynamics within a given spine (j) at delays determined by r_j . The intracellular calcium spike is represented by the quantity $[G_{ij}Y_{ij}]^+$, where

$$\frac{d}{dt}G_{ij} = \alpha_G(B_G - G_{ij})f_G(x_{ij} - \Gamma_G) - \beta_G G_{ij} \tag{11}$$

and

$$\frac{d}{dt}Y_{ij} = \alpha_Y(1-Y_{ij}) - \beta_Y[G_{ij}Y_{ij} - \Gamma_Y]^+ \tag{12}$$

In Equation 11, $f_G(x)$ is a step function:

$$f_G(x) = \begin{cases} 1 & (x > 0) \\ 0 & (x < 0) \end{cases} \tag{13}$$

In the brief interval when the calcium concentration at a particular spine exceeds a threshold activity Γ_S, CS-striosomal weight Z_{ij} at that particular spine becomes eligible for change that may be induced by dopaminergic bursts (N^+) or dips (N^-).

$$\frac{d}{dt}Z_{ij} = \alpha_Z[G_{ij}Y_{ij} - \Gamma_S]^+((A_Z - Z_{ij})N^+ - B_Z Z_{ij}N^-) \tag{14}$$

Third, the changes in the level of PPTN (P) are described by the following differential equations:

$$\frac{1}{\tau_{P1}}\frac{d}{dt}P_{pre-excite} = -P_{pre-excite} + (1 - P_{pre-excite})W_{SP}S \tag{15}$$

$$\frac{1}{\tau_{P2}}\frac{d}{dt}P_{pre-inhibit} = -P_{pre-inhibit} + (1 - P_{pre-inhibit})W_{SP}S \tag{16}$$

$$\frac{1}{\tau_P}\frac{d}{dt}P = background_P - P + (1-P)W_P input_{pre\_P} \tag{17}$$

where

$$input_{pre\_P} = \begin{cases} \left[P_{pre-excite} - P_{pre-inhibit} - \Gamma_{P12}\right]^+ & P_{pre-excite} > P_{pre-inhibit} \\ -\left[P_{pre-inhibit} - P_{pre-excite} - \Gamma_{P12}\right]^+ & P_{pre-excite} < P_{pre-inhibit} \end{cases} \tag{18}$$

$P_{pre-excite}$ and $P_{pre-inhibit}$ can be regarded as the effect of substance P and GABA on PPTN. Ventral striatum neurons can secrete substance P and GABA. Substance P excites the following neurons, while GABA inhibits the following neurons.

$input_{pre\_P}$ denotes the net effect of substance P and GABA. The author believe that this explanation is more realistic, but it needs more physiological experiments to be testified. The changes of the activity level of PPTN neurons depend on the background inputs, its decay and the net effect from the striatum.

Fourth, the changes in the level of ventral pallidum (VP) are described by the following differential equations:

$$\frac{1}{\tau_{VP1}}\frac{d}{dt}VP_{pre-excite} = -VP_{pre-excite} + (1 - VP_{pre-excite})W_{SVP}S \tag{19}$$

$$\frac{1}{\tau_{VP2}}\frac{d}{dt}VP_{pre-inhibit} = -VP_{pre-inhibit} + (1 - VP_{pre-inhibit})W_{SVP}S \tag{20}$$

$$\frac{1}{\tau_{VP}}\frac{d}{dt}VP = background_{VP} - VP + (1 - VP)W_{VP}input_{pre\_VP} \tag{21}$$

where

$$input_{pre\_VP} = \begin{cases} \left[VP_{pre-excite} - VP_{pre-inhibit} - \Gamma_{VP12}\right]^{+} & VP_{pre-excite} > VP_{pre-inhibit} \\ -\left[VP_{pre-inhibit} - VP_{pre-excite} - \Gamma_{VP12}\right]^{+} & VP_{pre-excite} < VP_{pre-inhibit} \end{cases} \tag{22}$$

The explanation is similar to Equations 15 ~ 18. The changes of the activity level of VP neurons result from the background inputs, its decay and the net effect from the striatum.

Fifth, changes in the level of GPb neurons are described by the following differential equation:

$$\frac{1}{\tau_{\mathrm{GPb}}}\frac{d}{dt}GPb = background_{GPb} - GPb + (1-GPb)(W_{SOG}\sum_{i,j}[G_{ij}Y_{ij} - \Gamma_S]^+ Z_{ij} - W_{VPG}VP)$$

(23)

The changes of the activity level of GPb neurons are determined by the background inputs, its decay, and the inhibitory effect from VP neurons and the disinhibitory input from striosome.

Sixth, changes in the level of LHb neural activity are described by the following differential equation:

$$\frac{1}{\tau_{LHb}}\frac{d}{dt}LHb = background_{LHb} - LHb + (1-LHb)W_{GL}[GPb - \Gamma_{GPb}]^+ \qquad (24)$$

The changes of the activity level of LHb neurons result from the background inputs, its decay and the excitatory input from the GPb.

Seventh, changes in the level of RMTg neurons are described by the following differential equation:

$$\frac{1}{\tau_{RMTg}}\frac{d}{dt}RMTg = background_{RMTg} - RMTg + (1-RMTg)W_{LR}[LHb - \Gamma_{LHb}]^+ \qquad (25)$$

The changes of the activity level of RMTg neurons depend on the background inputs, its decay and the excitatory input from the LHb.

Eighth, changes in the level of dopaminergic neurons (D) are described by the following differential equation:

$$\frac{1}{\tau_{\mathrm{D}}}\frac{d}{dt}D = background_D - D + (1-D)(W_{PD}[P-\Gamma_P]^+ - W_{RD}RMTg) - (D+h_D)\sum_{i,j}[G_{ij}Y_{ij} - \Gamma_S]^+ Z_{ij}$$

(26)

The changes of the activity level of dopaminergic neurons depend on the background inputs, its decay, the inhibitory effect from RMTg neurons and striosomes and the excitatory input from the PPTN.